Developing 'Updatable' Crash Prediction Model for Network Screening: A Case Study of Czech Two-Lane Rural Road Segments

Jiří Ambros (corresponding author)

CDV – Transport Research Centre (Centrum dopravního výzkumu, v.v.i.)

Líšeňská 33a, 636 00 Brno, Czech Republic

Tel. +420 541 641 362; Fax +420 541 641 712; E-mail jiri.ambros@cdv.cz

Veronika Valentová

CDV – Transport Research Centre (Centrum dopravního výzkumu, v.v.i.)

Líšeňská 33a, 636 00 Brno, Czech Republic

Tel. +420 541 641 355; Fax +420 541 641 712; E-mail veronika.valentova@cdv.cz

Jiří Sedoník

CDV – Transport Research Centre (Centrum dopravního výzkumu, v.v.i.)

Líšeňská 33a, 636 00 Brno, Czech Republic

Tel. +420 541 641 293; Fax +420 541 641 712; E-mail jiri.sedonik@cdv.cz

Word count: 4,300 (abstract, text, references) + 10 figures/tables × 250 = 6,800

TRR Paper No. 16-1817

Submission date 11/01/2015

ABSTRACT

The case study focuses on application of crash prediction models in network screening. The two main questions were (1) *What variables should be involved in the model?* and (2) *How long should the modeled time period be?* Answers to these questions should provide guidelines to developing ‘updatable’ crash prediction model, i.e. a model which is both reliable and simple, so that its updating for periodical network screening is not highly demanding.

To this end approximately 1,000 km (600 mi) of two-lane rural road network data from South Moravia (Czech Republic) was used. Based on 8 years of annual crash frequencies, together with exposure and geometrical variables, several variants of prediction models were developed. In order to study their quality, a series of consistency tests was applied, relative to comparison of models themselves, as well as their diagnostic performance.

As a result simple crash prediction models (including traffic volume, segment length and curvature change rate) were found as sufficient for network screening. Supposing that length and curvature are not likely to change often, only traffic volume data need to be periodically updated. Based on consistency analyses this time period should be 4 years. Under these conditions, models are currently being applied in the studied region; further planned activities include extensions to intersections and also to other Czech regions.

1 INTRODUCTION

Network screening (also referred to as 'identification of hazardous road locations' or 'hotspot identification') is defined as 'the process by which a road network is screened to identify sites that require safety investigation' (*1*). As the first step of road network safety management process, it is one of the most frequent tasks conducted by road agencies (*1*, *2*, *3*). According to recommended practices (*4*, *5*), network screening employs crash prediction models (also known as 'safety performance functions', SPFs) and empirical Bayes (EB) approach; in the end the list is produced which enables ranking the locations from most likely to least likely to realize a reduction in crash frequency with implementation of countermeasures (*3*).

Although the mentioned process has been known for several decades, it involves several decisions which have not often been described in the literature with sufficient detail. The main questions, addressed in this paper, are following:

1) *What variables should be involved in the model?* There are basically two types of models: simple and multivariate (*6*, *1*). The difference is in the independent variables (covariates): while simple model involves risk exposure only (i.e. traffic volume and segment length), multivariate models use also further variables, usually geometric characteristics. While increasing model complexity may provide better insight into modeled safety performance, it has often been found that additional predictors were not beneficial. For example studies of UK single carriageways (undivided two-lane roads), conducted in the 1990s (*7*), found that only a small fraction of explanatory variables significantly improved the model fit; also recent follow-up study (*8*) confirmed that the best fitting models did not include any of geometric features. Similar conclusions have been reached by Finnish researchers in the 1990s (*9*), and dictated their use of simple models since then (*10*). In recent US study (*11*), only a few variables were found to explain most of the variation in the crash data. Also previous Czech study (*12*) showed that simple models may be sufficient for network screening (however only one time period was used, with arbitrarily set length).
2) *How long should the modeled time period be?* A period between 1 and 5 years is usually recommended for network screening, with 3-year period being the most frequent (*4*). It is a compromise between the need for quick detection and the need for accumulating a sufficient crash numbers to permit analysis (*13*). On the other hand, longer time periods may cause problems with instability of conditions which may not reflect current traffic situation anymore (*14*). Probably due to these issues no specific guidelines for time period choice is usually provided; one exception was the simulation study of Cheng and Washington (*15*) which concluded there is little gain in the network screening accuracy when using a period longer than 6 years.

Both questions are interrelated and critical in the process of developing 'updatable' crash prediction model. Such 'updatable' model should be sufficiently reliable (describing safety performance of a modeled dataset) while also enough simple and parsimonious so that its updating (in a specific time period to be estimated) is not highly demanding.

This case study presents a development of ‘updatable’ crash prediction model for network screening in almost 1,000 km (approx. 600 mi) of two-lane rural road segments in South Moravian region, Czech Republic. It is meant as the first application of such study in Czech condition, aiming to prove the approach feasibility as well as practical applicability for the needs of a regional road agency. Based on 8 years of annual crash frequencies, together with exposure and geometrical variables (described in section 2), several variants of prediction models will be developed. In order to study their quality, a series of tests will be applied (section 3 describes comparison of models, section 4 is devoted to model performance). The discussion and conclusions (section 5) will provide answers to the two above stated questions.

2 DATA PREPARATION

Segmentation

The studied network consists of road sections (excluding intersections) of undivided two-lane paved rural roads. Their total length (995 km, i.e. 618 mi) was divided into homogeneous segments with respect to following variables: annual average daily traffic (AADT), speed limit reduction, road category, number of lanes, paved shoulder. A change of any of these variables marked the end of a segment and beginning of another one. Figure 1 illustrates the principle of this segmentation, following the work of Cafiso et al. (*16*).

In order to obtain segment lengths which will be practical for follow-up safety inspections, segments longer than 500 m (1640 ft) were divided into 250 m (820 ft) parts. After this segmentation most of the segments (78%) were 250 m long, 17% were longer and 5% were shorter. Their total number was 3,764.

Variables

These segments were assigned specific values of response variable (crash frequency) and explanatory variables (exposure data, road and traffic characteristics, context and environment variables) which represent safety-related features. Reported crash frequency data for the period 2007 – 2014 were obtained from the Czech Traffic Police. They include all injury crashes, i.e. with slight, severe or fatal personal consequences. After exclusion of intersection-related crashes, there were 2,219 crashes in total, with frequency between 0 and 18 within a segment. Further explanatory variables were added:

- Annual average daily traffic (AADT) and percentage of heavy goods vehicles (HGV) data were used to represent the crash exposure. These data were acquired from Czech Road and Motorway Directorate, based on the results of national traffic census 2010.
- The segment length, as the second risk exposure variable; its generation was described in the previous paragraph.

- Road and traffic characteristics data were also obtained from a Czech Road and Motorway Directorate database, reflecting the state in July 2010. Density of intersections with minor rural roads and density of roadside facilities were computed as frequencies divided by segment lengths. Some variables were not well represented across their range of values: regarding number of lanes, 95% cases were two-lane segments; regarding presence of speed limit reduction, 99% cases were without reduction, i.e. with 90 kph (approx. 56 mph) speed limit applied. These two variables were thus removed.

- Another variable was related to the level of quality of road pavement. It is defined according to collected data about cracks and potholes in five classes, from 1 (excellent) to 5 (wrecking). Data collected by Czech company PavEx Consulting in 2011 were used (*17*), where sections were classified according to the worst quality level presence; however only the list containing sections with levels 4 (unsatisfactory) and 5 (wrecking) was available to authors. Therefore for further analyses each segment was assigned the ratio of length with the pavement quality level 4 or 5; the value is between 0 and 1 (0 means quality better than 4 or 5; 1 means total length in these quality levels).

- Some of context and environment variables were also acquired from the Czech Road and Motorway Directorate database. The data for other ones were collected additionally: these were average curvature change rate (CCR), as a traditional alignment consistency indicator, and forest environment, which is linked to wet surface or game crashes (*18*). CCR was computed as sums of angles between vertices divided by the sum of their lengths (*19*). The information about continuous forest around the road segment was collected manually from on-line maps, which utilize the data of Czech Environmental Information Agency (CENIA).

All variables and their characteristics are summarized in Table 1.

Modeling

Prior to the modeling, variables inter-correlation was investigated. The highest significant correlation was found between road category and paved shoulder (almost 0.67). Since the difference in lane width for each road category is minimal (category 7.5 is by 0.25 m narrower than 9.5 and 11.5), the connection may be therefore caused by the relation between road category and paved shoulder, which is governed by Czech standards: wider roads tend to be equipped with paved shoulder. Since the amount of intercorrelation was considerable and given the fact that road category is a derived variable, it was removed from the data set.

A negative binomial regression modeling was used to develop the models (for details see e.g. *20*), with AADT modeled in function form $AADT^{\beta_1} \cdot \exp(\beta_2 \cdot AADT)$ (following *21*). The required prediction model form was then as follows:

$$P_i = \exp(\beta_0) \cdot AADT_i^{\beta_1} \cdot \exp(\textstyle\sum_{i=2}^{n} \beta_i x_i) \qquad (1)$$

where β_i are coefficients to be estimated in modeling and x_i are explanatory variables. For estimation SPSS procedure GENLIN was used.

3 COMPARISON OF MODELS

As indicated, depending on the number of explanatory variables, model type may be simple or multivariate. In order to study the suitable type within available time frame (8 years), following investigation scheme was planned:

– 3-year period is most often used in the literature – thus it was chosen as a minimum time period. Within 8-year time frame, there are 6 variants of 3-year periods.

– 5 years are usually seen as the maximum; in order to check the performance above this level, 6-year period was chosen as the maximal. Within 8-year time frame, there are 3 variants of 6-year periods.

– 4-year periods (5 variants) and 5-year periods (4 variants) were chosen in a similar way, using overlapping variants. For illustration see Figure 2.

Crash prediction models were developed for all 18 variants, in a backward elimination manner, with only the variables with at least 95% statistical significance being kept in the model.

Variants of 5-year models could not be built; the error (caused by singularity or nonpositive definiteness of Hessian matrix) is 'frustrating but common occurrence in applied quantitative research' (*22*). Therefore only model variants for 3-year, 4-year and 6-year time periods were developed. In order to see the effect of variable choice on the 'least common denominator', Figure 3 shows the resulting significance of explanatory variables for 3-year variants (as described in Figure 2). Gray cells indicate significance, white cells indicate lack of significance. For all the variants, three variables (in bold) were always statistically significant: logarithm of AADT, length and curvature change rate (CCR).

In the following text the two model types are distinguished:

– the ones, including the three mentioned variables (log AADT, length, CCR), will be referred to as *simple*
– the ones including also other variables will be referred to as *multivariate*

In order to compare the performance (goodness-of-fit) of simple and multivariate models, various indicators may be used. For example Oh et al. (*23*) used five different measures to assess the external validity (Pearson correlation coefficient between observed and predicted crash frequencies, mean prediction bias, mean absolute deviation, mean squared prediction error, mean squared error), while noting that they all should be considered jointly. For the sake of brevity, a single indicator was used here – proportion of systematic variation in the original crash dataset explained by the model ($\% SV$) (e.g. *24*, *25*, *26*). The indicator $\% SV$ was computed for all variants of 3-year, 4-year and 6-year models. The values were averaged for individual time periods. The results are shown in Figure 4: in gray for multivariate models, in white for simple models.

From Figure 4 it is evident that results are all around 60%, and relatively similar for both model types, with differences below 3%. It means that the explanatory variables, which are missing in simple models, provide only minor improvement of model performance.

4 COMPARISON OF MODEL PERFORMANCE

As stated in the beginning, the study focus is on network screening, based on empirical Bayes method with crash prediction models. In principle crash prediction models were used to obtain predicted crash frequency (P) for each segment (i). Empirical Bayes estimate of expected crash frequency (EB) was then calculated, using predicted crash frequency, reported crash frequency and length-dependent overdispersion parameter (*5*, *27*). Finally potential for safety improvement (PSI) was obtained as a difference between predicted crash frequency and EB estimate (*28*).

$$EB_i = w_i \cdot P_i + (1 - w_i) \cdot R_i \qquad (2)$$

$$w_i = \frac{k_i}{k_i + P_i} \qquad (3)$$

$$k_i = k \cdot L_i \qquad (4)$$

$$PSI_i = EB_i - P_i \qquad (5)$$

where

EB_i	EB estimate
w_i	weight
P_i	predicted crash frequency
R_i	reported crash frequency
k_i	overdispersion parameter
L_i	segment length
PSI_i	potential for safety improvement

Values of PSI were used for network screening. After their descending ranking, two lists of segment numbers were developed. Following previous studies, 1%, 3% and 5% upper tails (top parts of distribution) were further used in order to investigate the differences (1% = 38 segments, 3% = 113 segments, 5% = 188 segments).

The objective is to assess the implications of choice of different modeling time periods on the results of network screening (i.e. ranked lists of segment identification numbers). Assessment may be done in terms of ‘consistency’. In literature various consistency tests have been used. For example Miranda-Moreno et al. (*29*) applied percentage deviation and Spearman correlation coefficient to compare the performance of two ranking criteria. Cheng and Washington (*15*) utilized false positives, false negatives and the effects of crash history duration to compare three hot spot identification methods (HSID). In order to evaluate the diagnostic performance of five HSID techniques, Elvik (*30*) used two epidemiological criteria (sensitivity and specificity). Montella (*31*) employed 4 consistency tests (site consistency test, method consistency test, total rank differences test, total score test) to assess the performance of seven HSID methods.

This literature review shows that consistency criteria have mostly been used for comparison of different screening methods. In this study the principle was adapted in order to compare the effect of using screening based on models from different time periods. Therefore the original meaning of 'identification with different methods' is applied in the meaning of 'identification in different time period' (all with three different levels of upper tails). In total four criteria were selected from the mentioned studies. Their descriptions and definitions are summarized in Table 2; further details may be found in *30, 31, 32*. Note that in Test 4, segments identified as hazardous in maximal (8-year) time period were considered as positives.

The tests were applied on all 28 possible variation pairs of lists of segment numbers (15 pairs for 3-year models, 10 pairs for 4-year models, 3 pairs for 6-year models), for 3 upper tails (1%, 3%, 5%). The figures 5 – 8 show the test results, averaged for specific combinations of upper tail and time period.

It is to be noted that Test 2 (method consistency test) showed relative similarity between the lists based on simple and multivariate models – the agreement was 90% or above for all variants. For clarity the following test results are reported for simple models only.

Based on the test outputs, shown in Figures 5 to 8, and test criterion definitions in Table 2, following results may be stated:

- In Test 1 (site consistency test), optimal (maximal) values are reached with 4-year models (Figure 5).
- In Test 2 (method consistency test), 6-year models provide optimal (maximal) values (Figure 6). However in other studies, which used this test to compare different HSID methods, values below 50% were usually reached as maximal (46.9% in *31*, 47.3% in *32*, 46.2% in *33*, all values for 5% upper tails). Should these values present acceptable threshold, 4-year models would be sufficient, since their values are all above 50% overlap.
- In Test 3 (total rank differences test), minimal value is the optimum. According to the Figure 7, one may choose between 4-year and 6-year time period.
- In Test 4 (epidemiological diagnostic test), optimal (maximal) values are reached with 6-year models. In addition simple and multivariate models were compared; the Figure 8 illustrates that both model types perform the closest with 4-year models.

5 DISCUSSION AND CONCLUSIONS

The case study focuses on application of crash prediction models in network screening (also ‘identification of hazardous road locations’ or ‘hotspot identification’). The two main questions were (1) *What variables should be involved in the model?* and (2) *How long should the modeled time period be?* Answers to these questions should provide guidelines to developing ‘updatable’ crash prediction model, i.e. a model which is both reliable and simple, so that its updating for periodical network screening is not highly demanding.

Regarding model type choice, simple and multivariate models were distinguished. In terms of proportion of systematic variation explained, simple alternatives are almost comparable to multivariate models (Figure 4). In Test 2 (method consistency test) relative similarity between the lists based on simple and multivariate models was reached – the agreement was 90% or above for all variants. In addition the results of Test 4 (epidemiological diagnostic test) illustrate that both model types perform relatively close, with optimum at model variants from 4-year time period.

Regarding choice of time period, several consistency tests were applied. Most of the results indicate that 4-year period is acceptable for developing a crash prediction model – in terms of site consistency, method consistency, and total rank differences. Unfortunately, with exception of Test 2 (with results percents of overlap between ranked lists), it is difficult to relate the test results to other studies, due to their several differences. For example with Test 4, different authors considered different time periods to represent ‘true mean’: while 8-year period (as an available maximum) was used in the presented case study, shorter period (3 years) was used by the others (*32*).

To sum up the simple models include explanatory variables of traffic volume (AADT), segment length and curvature change rate (CCR). Supposing that length and curvature are not likely to change often, only one variable (AADT) needs to be updated. Network-wide AADT update should follow national traffic census, which takes place every 5 years; in the meantime AADT growth factors (*34*) may be used, in overlapping time period of 4 years. Currently 2011 – 2014 time period has been used; as soon as 2015 crash data become available in 2016, new model for 2012 – 2015 period will be calibrated.

Under these conditions, ‘updatable’ crash prediction model is currently applied in the studied region. It is hoped to be a beneficial tool for road network safety management in South Moravian region. Further planned activities include extensions to intersections and also to other Czech regions.

ACKNOWLEDGMENTS

The help of Jan Novák and Robert Zůvala with computations of consistency tests is appreciated. The study was conducted with support of the Czech Ministry of Interior’s research project No. VG20112015013 ‘Identification and treatment of high risk road spots and sections’ (IDEKO) and research infrastructure of ‘Transport R&D Centre’ (CZ.1.05/2.1.00/03.0064).

REFERENCES

1. Persaud, B. N. *Statistical Methods in Highway Safety Analysis: A Synthesis of Highway Practice*. NCHRP Synthesis 295. Transportation Research Board of the National Academies, Washington, D.C., 2001.
2. Hauer, E., J. Kononov, B. Allery, and M. S. Griffith. Screening the Road Network for Sites with Promise. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1784*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 27–32.
3. *Highway Safety Manual. First Edition.* American Association of State Highway and Transportation Officials (AASHTO), Washington, D.C., 2010.
4. Elvik, R. A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 1830–1835.
5. Hauer, E., D. W. Harwood, F. M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1784*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 126–131.
6. Hauer, E. *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon Press, Oxford, 1997.
7. Walmsley, D. A., I. Summersgill, and C. Binch. *Accidents on modern rural single-carriageway trunk roads*. TRL Report 336. Transport Research Laboratory, Crowthorne, 1998.
8. Wood, A. G., L. J. Mountain, R. D. Connors, M. J. Maher, and K. Ropkins. Updating outdated predictive accident models. *Accident Analysis and Prevention*, Vol. 55, 2013, pp. 54–66.
9. Peltola, H., R. Kulmala, and V.-P. Kallberg. Why use a complicated accident prediction model when a simple one is just as good? In *22nd PTRC Summer Annual Meeting*, Warwick, UK, September 12–16, 1994.
10. Peltola, H., R. Rajamäki, and J. Luoma. A tool for safety evaluations of road improvements. *Accident Analysis and Prevention*, Vol. 60, 2013, pp. 277–288.
11. Saha, D., P. Alluri, and A. Gan. Prioritizing Highway Safety Manual's crash prediction variables using boosted regression trees. *Accident Analysis and Prevention*, Vol. 79, 2015, pp. 133–144.
12. Ambros, J., V. Valentová, and Z. Janoška. Investigation of difference between network screening results based on multivariate and simple crash prediction models. In *Transportation Research Board 94th Annual Meeting*, Washington, D.C., January 11–15, 2015.
13. Elvik, R. *Assessment and applicability of road safety management evaluation tools: Current practice and state-of-the-art in Europe*. TØI report 1113/2010. Institute of Transport Economics, Oslo, 2010.
14. Hauer, E., and B. N. Persaud. Problem of identifying hazardous locations using accident data. In *Transportation Research Record: Journal of the Transportation Research Board*,

No. 975, Transportation Research Board of the National Academies, Washington, D.C., 1984, pp. 36–43.

15. Cheng, W., and S. Washington. Experimental evaluation of hotspot identification methods. *Accident Analysis and Prevention*, Vol. 37, 2005, pp. 870–881.
16. Cafiso, S., A. Di Graziano, G. Di Silvestro, G. La Cava, and B. Persaud. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. *Accident Analysis and Prevention*, Vol. 42, 2010, pp. 1072–1079.
17. Gregor, J., and L. Mališ. *Stav povrchu vozovek silnic II. a III. třídy v Jihomoravském kraji k 30.4.2011*. PavEx Consulting, s.r.o., Brno, 2011.
18. Elvik, R., A. Høye, T. Vaa, and M. Sørensen. *The Handbook of Road Safety Measures, Second Edition*. Emerald, Bingley, 2009.
19. Dantas, A., M. Fowler, and G. Koorey. Effect of road network bendiness on traffic crash occurrence. In *IPENZ Transportation Group Technical Conference*, Tauranga, New Zealand, October 7–10, 2007.
20. Hauer, E. Statistical Road Safety Modeling. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1897*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 81–87.
21. Hauer, E., F. M. Council, and Y. Mohammedshah. Safety Models for Urban Four-Lane Undivided Road Segments. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1897*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 96–105.
22. Gill, J., and G. King. What to Do When Your Hessian Is Not Invertible: Alternatives to Model Respecification in Nonlinear Estimation. *Sociological Methods and Research*, Vol. 33, 2004, pp. 54–87.
23. Oh, J., C. Lyon, S. Washington, B. Persaud, and J. Bared. Validation of FHWA Crash Models for Rural Intersections: Lessons Learned. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1840*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 41–49.
24. Kulmala, R. *Safety at rural three- and four-arm junctions: Development and application of accident prediction models*. Publication No. 233. VTT Technical Research Centre of Finland, Espoo, 1995.
25. Vogt, A., and J. G. Bared. *Accident Models for Two-Lane Rural Roads: Segments and Intersections*, Report FHWA-RD-98-133. Federal Highway Administration (FHWA), McLean, 1998.
26. Greibe, P. Accident prediction models for urban roads. *Accident Analysis and Prevention*, Vol. 35, 2003, pp. 273–285.
27. Hauer, E. Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. *Accident Analysis and Prevention*, Vol. 33, 2001, pp. 799–808.
28. Persaud, B., C. Lyon, and T. Nguyen. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1665*, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 7–12.

29. Miranda-Moreno, L. F., L. Fu, F. F. Saccomanno, and A. Labbe. Alternative Risk Models for Ranking Locations for Safety Improvement. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 1–8.
30. Elvik, R. Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2083*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 72–75.
31. Montella, A. A comparative analysis of hotspot identification methods. *Accident Analysis and Prevention*, Vol. 42, 2010, pp. 571–581.
32. Cheng, W., and S. Washington. New Criteria for Evaluating Methods of Identifying Hot Spots. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2083*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 76–85.
33. Yu, H., P. Liu, J. Chen, and H. Wang. Comparative analysis of the spatial analysis methods for hotspot identification. *Accident Analysis and Prevention*, Vol. 66, 2014, pp. 80–88.
34. Bartoš, L., A. Richtr, J. Martolos, and M. Hála. *Prognóza intenzit automobilové dopravy, II. vydání*. Technical guidelines No. 225. EDIP s.r.o., Plzeň, 2012.

LIST OF FIGURES

LIST OF TABLES

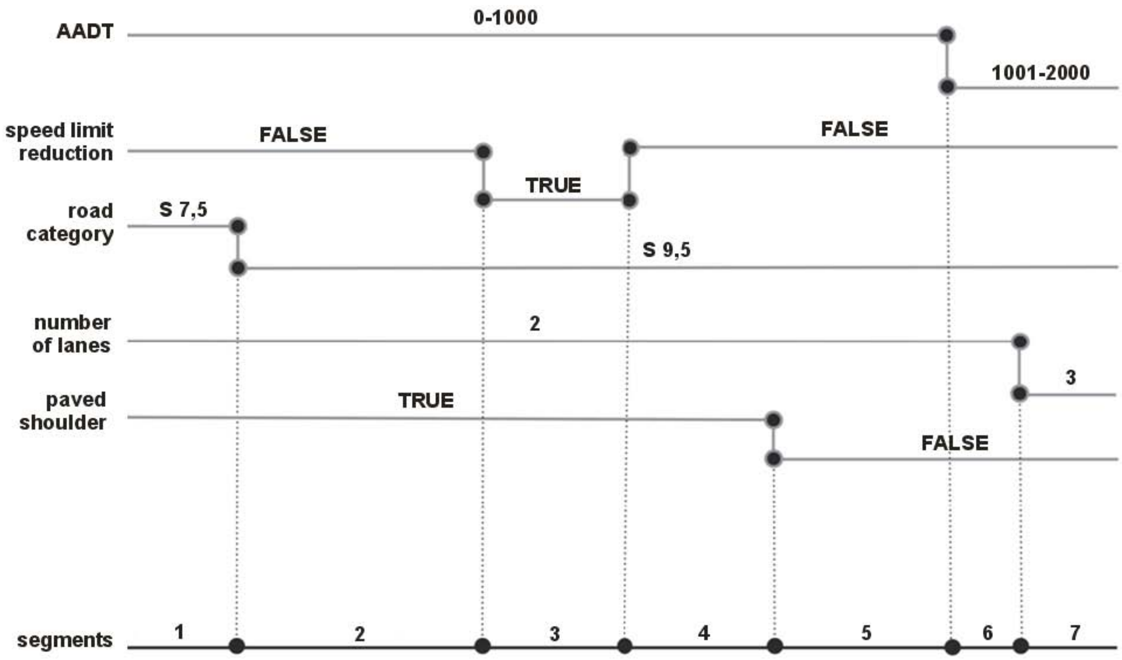

FIGURE 1 Principle of division into homogeneous segments.

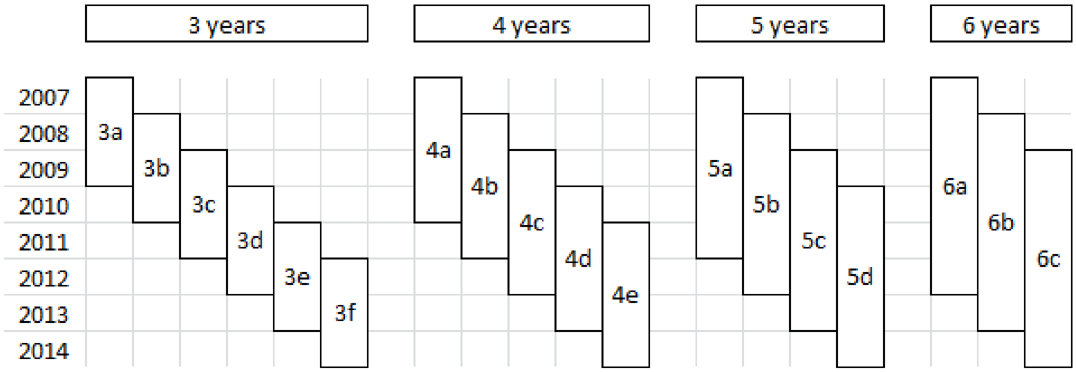

FIGURE 2 Scheme of investigated variants of time periods.

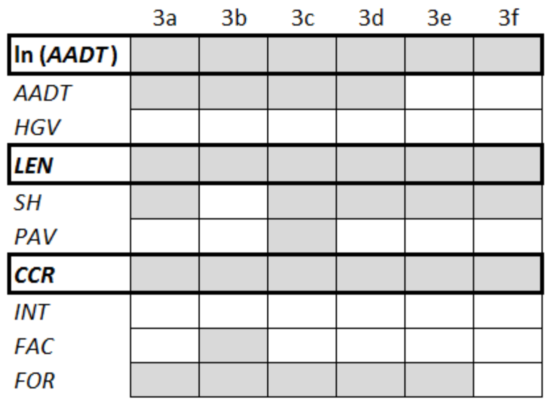

FIGURE 3 Overview of achieved statistical significance of explanatory variables in 3-year models (gray cells = significant, white cells = non-significant).

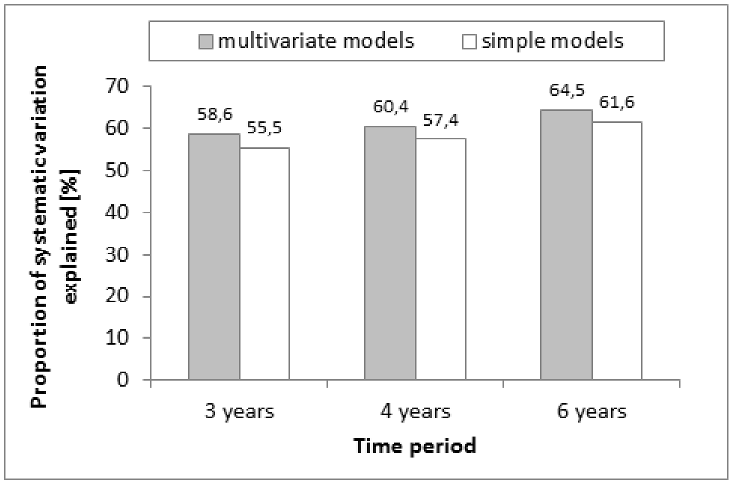

FIGURE 4 Comparison of average 'proportion of systematic variation explained' of multivariate and simple models.

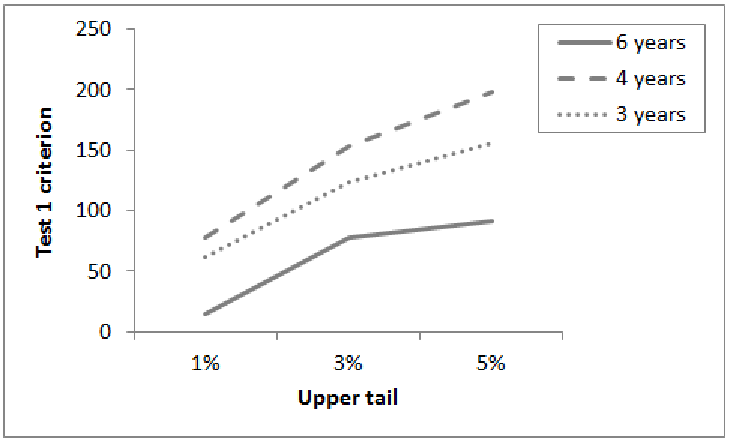

FIGURE 5 Results of Test 1 (site consistency test) for different time periods and upper tails.

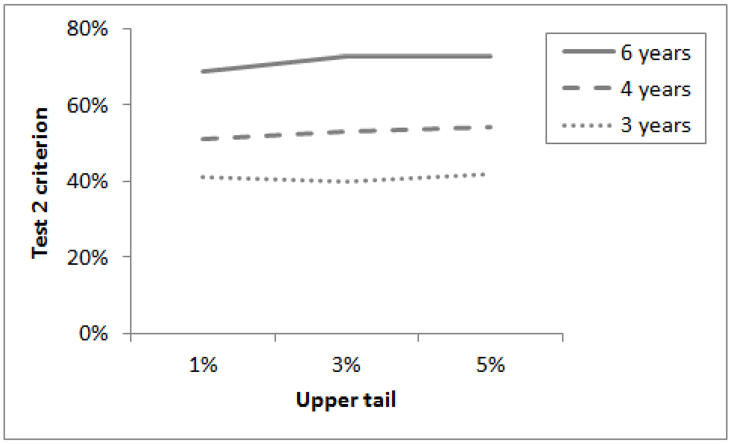

FIGURE 6 Results of Test 2 (method consistency test) for different time periods and upper tails.

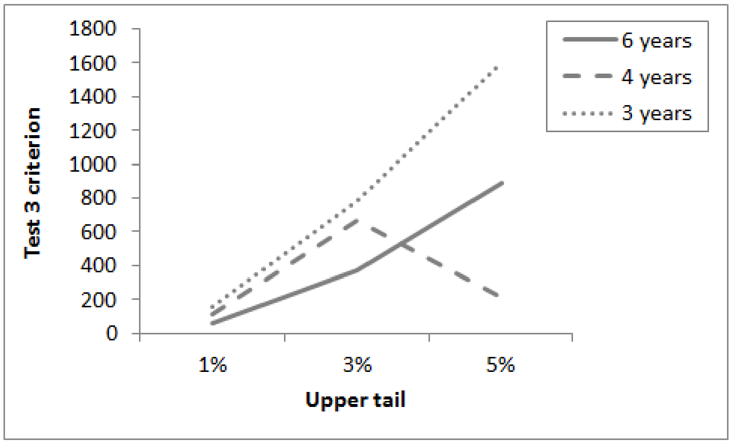

FIGURE 7 Results of Test 3 (total rank differences test) for different time periods and upper tails.

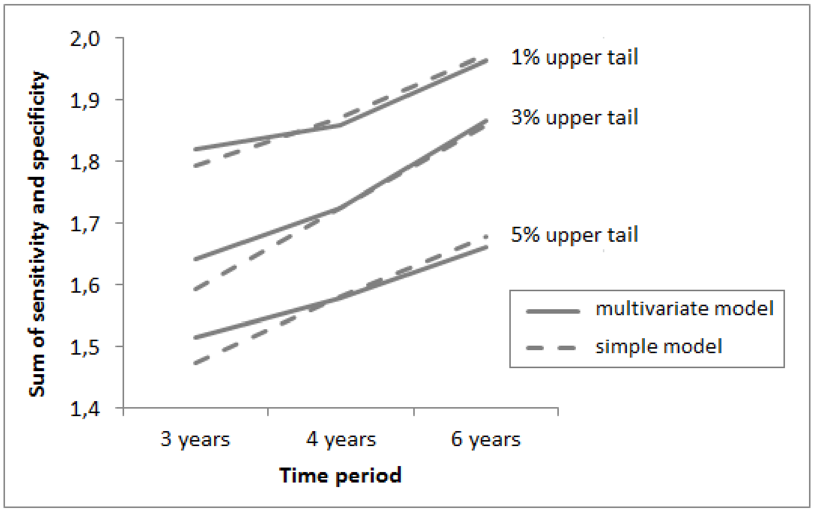

FIGURE 8 Results of Test 4 (epidemiological diagnostic test) for different time periods and upper tails.

TABLE 1 Overview of Data with Description and Descriptive Statistics of Variables

Data type	Abbr.	Description	Data type and unit*	Descriptive statistics (min / max / mean / SD or frequencies)
Crash data	R	8-year frequency of reported injury crashes	count	0 / 18 / 0.59 / 1.21
Exposure data	$AADT$	Annual average daily traffic	continuous [vehicle per day]	91 / 18,498 / 2,459.37 / 2,229.86
	HGV	HGV percentage	continuous	0.06 / 0.50 / 0.18 / 0.06
	LEN	Segment length	continuous [m]	51.00 / 499.88 / 264.29 / 64.03
Road and traffic data	CAT	Road category	binary (0 = 7.5 m; 1 = 9.5 or 11.5 m wide)	0: 3,156; 1: 608
	SH	Paved shoulder	binary (0 = FALSE; 1 = TRUE)	0: 3,333; 1: 431
	PAV	Pavement quality	continuous	0 / 1 / 0.49 / 0.49
Context and environment data	CCR	Average curvature change rate	continuous [gon per km]	0.0 / 1,498.18 / 98.03 / 134.27
	INT	Density of intersections with minor rural roads	continuous [number per km]	0.0 / 16.90 / 1.16 / 2.40
	FAC	Density of roadside facilities	continuous [number per km]	0.00 / 52.00 / 2.58 / 5.76
	FOR	Forest environment	binary (0 = FALSE; 1 = TRUE)	0: 2,977; 1: 787

* Note: 1 m (meter) = 3.3 ft, 1 km (kilometer) = 0.6 mi

TABLE 2 Summary and Description of Applied Consistency Tests

Test	Test premise	Test criterion
Site consistency test (Test 1)	A segment identified as risky during time period i should also be identified as risky in time period $i + 1$.	The method that identifies segments in time period i, with the highest crash frequency in a time period $i + 1$, is the most consistent.
Method consistency test (Test 2)	A list of segments, which were identified as risky during time period i, should be similar to the list of segments, which will be identified as risky in time period $i + 1$.	The method that identifies segments in time period i, whose list has the largest overlap with the list of segments identified in time period $i + 1$, is the most consistent.
Total rank differences test (Test 3)	The segments in the lists, produced in Test 2, should have similar rankings between time periods i and $i + 1$.	The method with the smallest sum of total rank differences between the lists from time periods i and $i + 1$ is the most consistent.
Epidemiological diagnostic test (Test 4)	The method should identify as many of the truly risky segments as possible (sensitivity), and as few of truly non-risky segments as possible (specificity).	The method with the greatest sum of sensitivity and specificity is the most consistent.